

RECOMBINANT WATERMELON (*Citrullus lanatus*) HYDROPEROXIDE LYASE AND USES THEREOF

Field of the Invention

This invention relates to the production of a hydroperoxide lyase (HL) protein in host cells *via* recombinant expression of the protein. Recombinant HL protein, DNA sequences encoding the protein, vectors containing these DNA sequences and hosts containing the vectors are provided, together with methods for recombinantly producing such protein, DNA sequences, vectors and hosts.

Background of the Invention

Fatty acid hydroperoxide lyase (HL) is an enzyme of the octadecanoid pathway that cleaves a C-C bond in the hydroperoxides of polyunsaturated fatty acids to generate aldehydes and ω -oxo-acids. This enzyme is widely distributed in plants and is involved in the biosynthesis of volatile aldehydes and alcohols. Its presence in plants can be recognized by the fact that crushed leaves of nearly every plant generate a grassy smell that is characteristic of C6-aldehydes. The short-chain volatile aldehydes and their reduced derivatives, alcohols, are important constituents of the characteristic flavors of fruits, vegetables and green leaves.

The six-carbon volatile compounds, leaf aldehyde [(2E)-hexanal] and leaf alcohol [(3Z)-hexanol], are important components of the aroma and flavor of fruits and vegetables and are associated with 'green notes' of leaves. These compounds are widely used as flavors in foods and beverages and some of the C6 aldehydes are also reported to have anti-microbial properties. These volatile compounds are produced by fatty acid hydroperoxide lyase (HL) by cleaving 13-hydroperoxy fatty acids, mainly from linolenic and linoleic acids.

Commercial production of natural 'green note' compounds is generally achieved by fractional distillation of essential oils such as mint oil or by the combined action of lipoxygenase and hydroperoxide lyase on unsaturated fatty acids using plant material from different sources. These processes provide low yields and are dependant on specific plant materials. Therefore, there is a need for improved methods of generating 'green note' compounds.

Summary of the Invention

It has now been found that high reproducible yields of "green note" compounds (e.g.,[(2E)-hexanal) can be obtained independent of plant materials and in the absence of unwanted side reaction (e.g. isomerase activity) by transfer of the gene coding for HL from plant into host cells, subsequent expression of the gene, addition of linolenic acid as substrate and lipoxygenase (or soybean as the enzyme source) to produce the leaf aldehyde [(2E)-hexanal].

Thus, in a first aspect of this invention, there are provided isolated DNA sequences encoding a protein having the amino acid sequence of SEQ ID NO. 1. In a preferred embodiment the isolated nucleic acid encodes a polypeptide having at least about 80% amino acid sequence similarity to SEQ ID NO. 1 and having HL activity.

In another aspect of the invention, there is provided an expression vector comprising a nucleic acid molecule encoding a polypeptide having at least about 90% sequence similarity to SEQ ID NO. 1 and having HL activity.

In yet another aspect of the invention there is provided a host cell transformed with an expression vector comprising a nucleic acid molecule encoding a polypeptide having at least about 90% sequence similarity to SEQ ID NO. 1 and having HL activity. In a preferred embodiment the host cell is *E. coli*.

In another aspect of the invention there is provided a method of making hydroperoxide lyase comprising culturing a host cell transformed with an expression vector comprising a nucleic acid molecule encoding a polypeptide having at least about 90% sequence similarity to SEQ ID NO. 1 and having HL activity under conditions that enable expression of the nucleic acid, and isolating the polypeptide from the host cells.

The invention also provides transgenic plants comprising a recombinant nucleic acid molecule comprising a nucleic acid molecule encoding a polypeptide having at least about 90% sequence similarity to SEQ ID NO. 1 and having HL activity.

In another aspect of the invention there is provided a method for producing green note compounds comprising reacting fatty acid hydroperoxide in the presence of isolated recombinant protein having the amino acid sequence of SEQ ID NO. 1 or a polypeptide having at least about 90% sequence similarity to SEQ ID NO. 1 and having HL activity; and reacting the resulting aliphatic aldehydes in the presence of isomerase and/or alcohol dehydrogenase.

Brief Description of the Drawings

Figure 1 is a schematic drawing of the hydroperoxide lyase pathway.

Figure 2 is the amino acid sequence of the watermelon (*Citrullus lanatus*) hydroperoxide lyase (SEQ ID NO. 1).

Figure 3 is the nucleotide sequence of the watermelon (*Citrullus lanatus*) hydroperoxide lyase gene (SEQ ID NO. 2).

Figure 4 is a bar graph showing the activity of HL over-expressed in tobacco plants. N lines contain watermelon HL; WT is a non-transformed tobacco plant line; and A24 is the highest expresser of *Arabidopsis* HL.

Figure 5 is a bar graph showing C6 aldehyde formation from HL over-expressed in tobacco leaves in comparison to normal (non-transformed) watermelon leaves. C1HL-Ni6: watermelon HL transgenic tobacco; AtHL-N24: *Arabidopsis* HL transgenic tobacco.

Detailed Description of the Invention

Fatty acid hydroperoxide lyase (HL) has been cloned from a variety of sources. The gene was first cloned from the bell pepper fruit. (Matsui et al., 1996). Moxon et al. reported in 1997 that the product of the cDNA associated with an *Arabidopsis* EST sequence 94J16 has HL activity when expressed in *E. coli*. Both Bate et al. (1998) and Matsui et al., (1999) have since confirmed that this *Arabidopsis* cDNA encodes an HL. Many additional HLs have been identified and cloned from several plant species, including tomatoes and potatoes. However, until now, the gene has not been isolated from watermelon, whose leaves are reported to be among the highest C6-aldehyde producers among plant tissues.

A watermelon (*Citrullus lanatus*) HL cDNA was cloned via RT-PCR. For cloning the gene, degenerate primers were designed based on conserved sequences of published sequences of several HL and AOS genes. The full sequence of the HL cDNA obtained using the primers was determined using RACE. The nucleotide sequence of the *Citrullus lanatus* HL gene is shown in Figure 3, while the derived amino acid sequence is shown in Figure 2. The watermelon polypeptide sequence has highest sequence similarity to cucumber 13-HL and 63, 59, 59, 58, 57 and 47% sequence identity with HL from guava, tomato, *Nicotiana attenuate*, potato, alfalfa, green pepper, *Arabidopsis* and barley, respectively.

As used herein the term "substantially homologous," means that a particular sequence, for example, a mutant sequence, varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. For purposes of the present invention, DNA sequences having at least 90 percent homology (or sequence similarity), encoding equivalent biological properties, and showing equivalent expression characteristics are considered substantially homologous. Sequences having lesser degrees of homology, encoding comparable HL activity, and showing equivalent expression characteristics, *e.g.*, fragments of the nucleotide sequence SEQ ID No:2 are considered substantial equivalents. Generally, homologous DNA sequences can be identified by cross-hybridization under standard hybridization conditions of moderate stringency.

The present invention also provides vectors and expression vectors containing the DNA sequences of the present invention, hosts containing such vectors for the production of proteins with HL activity, and processes for the production of such DNA sequences, recombinant vectors and host cells.

The invention also provides recombinant proteins with HL activity. Specifically a protein with HL activity is defined to include a polypeptide having the amino acid sequence SEQ ID No:2 or any protein or polypeptide having an amino acid sequence which is substantially homologous to the amino acid sequence SEQ ID No:2 and further having the following biological activity: When the protein or polypeptide is incubated under suitable conditions and a suitable amount of substrate such as 13-(S)-linolenic acid hydroperoxide is added, the formation of (2E)-hexenal is observed.

As used herein the term recombinant proteins having at least about 90% sequence

similarity to SEQ ID NO. 1 and having HL activity includes proteins modified deliberately, as for example, by addition of specific sequences that preferably bind to an affinity carrier material. Examples of such sequences are sequences containing at least two adjacent histidine residues (see in this respect European Patent No. 282 042). Such sequences bind selectively to nitrilotriacetic acid nickel chelate resins (Hochuli and Dobeli, Biol. Chem. Hoope-Seyler 368, 748 (1987); European Patent No. 253 303). Proteins with HL activity which contain such a specific sequence can, therefore, be separated selectively from the remaining polypeptides. The specific sequence can be linked either to the C-terminus or the N-terminus of the proteins with HL activity.

Methods for the expression, isolation and purification of watermelon HL proteins and proteins having at least about 90% sequence similarity to SEQ ID NO. 1 and having HL activity are also provided. For example, recombinant HL protein can be prepared by cloning of DNA sequences encoding proteins with HL activity. DNA sequences encoding proteins with HL activity can be cloned using a variety of techniques. Using the methods described in this application cDNAs encoding proteins with HL activity or fragments thereof can be produced. These cDNAs can be isolated and amplified by PCR technique using oligodeoxynucleotide DNA primers by conventional techniques.

The cDNA (SEQ ID No:2) encoding the amino acid sequence SEQ ID No:1 was obtained using the DNA primers described in the examples. By using conventional techniques, this cDNA has been isolated from RT-PCR products using RNA derived from watermelon (*Citrullus lanatus*) leaves. A RT-PCR band of an expected size was obtained from watermelon leaf total RNA using a degenerate primer pair. The full length watermelon HL cDNA was isolated and cloned into a CMV 35S-driven binary vector.

The cDNA may be obtained not only from RT-products, but by other conventional techniques, e.g., by cloning genomic DNA, or fragments thereof, purified from the desired cells or plant tissues. These procedures are described by Sambrook et al., in "DNA Cloning: A Practical Approach", Vol. I and II, D. N. Glover, ed., 1985, MRL Press, Ltd., Oxford, U. K; Benton and Davis, Science 196, 180-182 (1977); and Grunstein and Hogness, Proc. Nat. Acad. Sci. 72, 3961-3965 (1975).

To obtain cDNA encoding watermelon proteins with HL activity cDNA libraries obtained using RNA isolated from watermelon leaf may be screened by conventional DNA hybridization techniques by the methods of Benton and Davis, supra, or Grunstein and Hogness, supra, using labeled HL gene fragments, for example. Clones which hybridize to the labeled gene fragments are analyzed, e.g., by restriction endonuclease cleavage or agarose gel electrophoresis. After isolating several positive clones the positive insert of one clone is subcloned, e.g., into phagemids, and sequenced by conventional techniques.

Clones derived from genomic DNA may contain regulatory and intron DNA regions in addition to coding regions; clones derived from cDNA will not contain intron sequences. In the molecular cloning of the gene from genomic DNA, DNA fragments are generated, some of which will encode the desired gene. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNase in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The linear DNA fragments can then be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography.

Whatever the source, the DNA sequence encoding proteins with HL activity may be cloned into a suitable vector for propagation of the DNA by methods known in the art. Any

commercially available vector may be used. For example, the DNA may be inserted into a CMV-S35 promoter-driven binary vector. Appropriate vectors for use with bacterial hosts are described by Pouwels et al., in "Cloning Vectors: A Laboratory Manual", 1985, Elsevier, N.Y. As a representative but nonlimiting example, useful cloning vectors for bacterial use can comprise a selectable marker and a bacterial origin of replication derived from commercially available plasmids which are in turn derived from the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and pGEM1 (Promega Biotec, Madison, Wis., USA).

The DNA sequences encoding proteins with HL activity inserted in these commercially available vectors can be verified by methods known in the art, *e.g.*, by standard nucleotide sequencing techniques.

DNA sequences that code for proteins with HL activity from plants other than watermelon may be used herein. Accordingly, while specific DNA has been cloned and sequenced in relation to the DNA sequence in watermelon leaves, any plant cell or tissue potentially can be used as the nucleic acid source of the protein with HL activity. DNA sequences having at least about 90% sequence similarity to SEQ ID NO. 2 are then selected for use in the present invention.

Cloned DNA sequences that code for watermelon proteins having HL activity or DNA sequences that encode a polypeptide having at least about 90% sequence similarity to SEQ ID NO. 1 and having HL activity can be expressed in hosts to enable the production of these proteins with greater efficiency. Techniques for these genetic manipulations are specific for the different available hosts and are known in the art.

For expression of proteins with HL activity in hosts, in principle, all vectors which

replicate and express DNA sequences encoding the proteins with HL activity in the chosen host are suitable. Expression vectors suitable for use in prokaryotic host cells are mentioned, for example, in the textbook "Molecular Cloning--A Laboratory Manual", Cold Spring Harbor Laboratory (1982), of Maniatis et al. Examples of other vectors are plasmids of the pDS family (Bujard et al., Methods in Enzymology, eds. Wu and Grossmann, Academic Press, Inc., Vol. 155, 416-433 (1987)).

Such prokaryotic expression vectors which contain the DNA sequences coding for proteins with HL activity operatively linked with an expression control sequence can be incorporated using conventional methods into any suitable prokaryotic host cell. The selection of a suitable prokaryotic host cell is determined by different factors which are well-known in the art.

Suitable prokaryotic host organisms include gram-negative and gram-positive bacteria, for example *E. coli* and *B. subtilis* strains. Examples of prokaryotic host organisms are *E. coli* strain M15 (described as strain OZ 291 by Villarejo et al. in J. Bacteriol. 120, 466-474 [1974] and *E. coli* W3110 [ATCC No. 27325]). In addition to the aforementioned *E. coli* strains, however, other generally accessible *E. coli* strains such as *E. coli* 294 (ATCC No. 31446) and *E. coli* RR1 (ATCC No. 31343) can also be used.

Plants can also be used as hosts for the recombinant production of proteins with HL activity. Transfer of the gene coding for the protein with HL activity may be achieved by a variety of methods (for review see Potrykus and Spangenberg, eds., Gene transfer to plants. A laboratory manual, Springer Verlag, Heidelberg, Germany (1995)), whereby the HL gene is integrated into the chromosome of the host plant. Homologous expression--overexpression--of the protein with HL activity can be achieved, for example, by transforming a watermelon plant

with the nucleic acid of SEQ ID NO. 2 or a nucleic acid encoding a polypeptide having at least about 90% sequence similarity to SEQ ID NO. 1 and having HL activity. Other examples for plant hosts for the production of recombinant HL protein include, but are not limited to maize (*Zea mays*, Ishida et al., Nature Biotechnology 14, 745-750 (1996)), flax (*Linum usitatissimum*, Dong and Mchughen, Plant Sci. 88 (1), 61-71 (1993)) and soybean (*Glycine max*, Christou et al., Tibtech 8, 145-151 (1990)).

Transgenic plants of the invention can be used as a source of green note compounds, for example, by homogenizing plant tissues or biomass with linoleic acid and lipoxygenase or soybean as an enzyme source. Also, the nucleic acids of the invention can be expressed in microorganisms, such as bacteria and yeast, and their biomass can be used in the same way to produce green note compounds. The green note compounds can be isolated from the transformed organisms by any process known to those of skill in the art.

For the isolation of small amounts of proteins with HL activity expressed in prokaryotic host cells for analytical purposes, e.g., for polyacrylamide gel electrophoresis, the host cells can be disrupted by treatment with a detergent, e.g., sodium dodecyl sulphate (SDS). Larger quantities of the HPO lyase protein can be obtained by mechanical [Charm et al., Meth. Enzymol. 22, 476-556 (1971)], enzymatic (lysozyme treatment) or chemical (detergent treatment, urea or guanidinium hydrochloride treatment, etc.) treatments followed by use of known methods, e.g., by centrifugation at different gravities, precipitation with ammonium sulphate, dialysis (at normal pressure or at reduced pressure), preparative isoelectric focusing, preparative gel electrophoresis or by various chromatographic methods such as gel filtration, high performance liquid chromatography (HPLC), ion exchange chromatography, reverse phase chromatography and affinity chromatography (e.g., on Sepharose.RTM. Blue CL-6B).

The proteins of the invention that have HL activity can be used in the production of natural "green note" compounds by catalyzing the formation of aldehydes from fatty acid hydroperoxide. The term "green note" compounds relates to leaf aliphatic aldehydes and leaf aliphatic alcohols, e.g., cis-3-hexenol and trans-2-hexenal. Hence the present invention also provides a process for the production of natural "green note" compounds, which process comprises the steps of:

- a) reacting fatty acid hydroperoxide in the presence of a protein having the amino acid sequence of SEQ ID NO. 1 or a polypeptide having at least about 90% sequence identity to SEQ ID NO. 1 and having HL activity; and
- b) reacting the resulting aliphatic aldehydes in the presence of isomerase and/or alcohol dehydrogenase.

In the process for the production of "green note" compounds the watermelon HL protein can be used in isolated form, or alternatively, in form of cell-free extracts obtained from host cells containing vectors for the production of protein with HL activity. In a preferred embodiment of the invention, the process for the production of natural "green note" compounds is employed to produce cis-3-hexenol. A preferred process for producing cis-3-hexenol comprises the steps of:

- a) reacting 13-(S)-hydroperoxide linolenic acid and/or hydrolyzed linseed oil in the presence of recombinant proteins of the invention having HL activity; and
- b) reducing the resulting cis-3-hexenal with alcohol dehydrogenase.

The green note compounds, *e.g.* cis-3-hexenol, prepared by the process of the present invention can be used as odorant and/or flavorant and worked into odorant and/or flavorant compositions in a manner known in the art.

Having now generally described this invention, the same will become better understood by reference to the specific examples, which are included herein for purpose of illustration only and are not intended to be limiting unless otherwise specified.

Example 1

Cloning of full length cDNA of watermelon hydroperoxide lyase

A pair of degenerate primers, GGN TTY AAYAGCN TWY GGN GG (upstream primer) (SEQ ID NO. 3) and CYT TNG CNG AYT GYT GYT TRT T (downstream primer) (SEQ ID NO. 4) was designed from published HL sequences of several plants and synthesized. Total RNA was isolated from watermelon leaf tissues using the trizol® reagent (Life Technologies) and the protocol suggested by the manufacturer. RT-PCR was conducted using the total RNA as template and a RT-PCR band of ~500 base pairs, which is the expected size for a HL protein was obtained. The RT-PCR product was partially sequenced, demonstrating that the RT-PCR product contained a portion of the coding region of watermelon HL. The RT-PCR product was partially sequenced, demonstrating that the RT-PCR product contained a portion of the coding region of watermelon HL.

Sequencing of the full length cDNA was accomplished using Clontech's SMART RACE kit. Two primers were designed from the partial sequence obtained:

Forward primer: CCG GCT CCG GTC TGA CAT TCG AGT CGG (SEQ ID NO.5)

Reverse primer: GCT CGC TCG GTA GTC CCG TCT GCG GCC CG (SEQ ID NO. 6)

With these gene-specific primers and primers supplied by the kit, the sequence of the coding region of the watermelon HL was obtained.

From the sequence obtained, the following primer pair was designed for cloning the full length cDNA of watermelon HL:

5'-end primer: CGC ACT AGT ATG AAG GTC ACC ATG ACC TC

3'-end primer: GGT AAG CTT CAG TTG GTC CTT TGA AAA GC

The 5'-end primer was designed to contain the SpeI-recognition site just before the start codon, while the 3'-end primer contained the HindIII recognition site just after the stop codon. The PCR product obtained using this primer pair was ligated into cloning vector pGEM-T Easy (Promega) by TA ligation.

Example 2

Expression of watermelon hydroperoxide lyase cDNA in plants

The isolated cDNA was then digested with SpeI and ligated into a CMV 35S promoter-driven binary vector pKYLX71:35S2 digested with XbaI, which produces a compatible 5'-extension. The binary vector was transferred into *Agrobacterium tumefaciens* strain GV3850 using freeze-thaw method and its integrity was confirmed by PCR. Using *Agrobacterium* containing the watermelon HL, tobacco plants were transformed by a leaf disc method, and the transgenic lines were selected for antibiotic resistance. The HL activity of the tobacco leaf tissues was analyzed and confirmed that transformation and expression of the watermelon HL cDNA had occurred. The enzyme extracts obtained from leaf tissues were reacted with 13-

hydroperoxy linolenic acid in the presence of yeast alcohol dehydrogenase (ADH) and NADH, where the production of aldehydes was measured by the reduction of NAD consumed by ADH converting aldehyde to alcohol spectrophotometrically. The products formed from 13-hydroperoxy linolenic acid in the leaf tissues of the transformed tobacco leaf were confirmed to be (3Z)-hexenal and (2E)-hexenal by GC-MS.

Example 3

Comparison of hydroperoxide lyase activity in transgenic plants

To compare the watermelon HL activity with another HL gene, HL cDNA was isolated from Arabidopsis and transgenic tobacco plants were generated following the same procedure described in example 2. HL activity was measured spectrophotometrically as described in example 2. The results shown in Figure 4 demonstrated the much higher activity of watermelon HL than Arabidopsis HL.

Example 4

Comparison of C₆ aldehyde production from transgenic plants

Twenty mg of leaf tissues of transgenic tobacco plants over-expressing watermelon HL or Arabidopsis HL or watermelon were homogenized with 5 mg of linolenic acid (Sigma) and 25 mg of soybean acetone powder as a lipoxygenase source in 2.0 mL of water at room temperature for 3 minutes. 0.224 mg (2.0 micromole) of (2E)-heptenal was added as an internal standard. The reaction products were extracted with 1 mL of pentane and analyzed with GC-MS. The results shown in Figure 5 demonstrated the higher production rate of (3Z)-hexenal by watermelon HL expressed in tobacco plants than Arabidopsis HL as well as watermelon leaves.

Example 5

Comparison of hydroperoxide lyases *C. lanatus*, *Arabidopsis thaliana* and *N. tabacum* expressed in *E. coli*

To compare the relative activities HLs from *C. lanatus* (CIHL), *Arabidopsis* (AtHL) and *N. tabacum* (NtHL), we produced these proteins in *E. coli* and measured HL enzyme activity in crude extracts of the *E. coli* cultures. Table I demonstrates that these clones all possess HL enzyme activity when compared to the vector control. Activity for the full-length CIHL protein was considerably higher (approximately 250-fold or more) compared to NtHL and AtHL.

Table I: HL activity of crude *E. coli* extracts containing (His)₆-tagged protein or (His)₆-tagged fusion protein

<i>E. coli</i> lysates expressing the following constructs	HL Activity (nmol/mg protein/min)
vector control	9.02 ± 2
CIHL-3-3 lysate	2,379 ± 516
NtHL-F6-3 lysate	28 ± 2.4
AtHL-F5-14 lysate	57

Values represent means ± SE of replicates (n=3-6) except the AtHL for which we only have one replicate so far.